

# UNIVERSAL PARTIAL WORDS OVER NON-BINARY ALPHABETS

BENNET GOECKNER, CORBIN GROOTHUIS, CYRUS HETTLE, BRIAN KELL,  
PAM KIRKPATRICK, RACHEL KIRSCH, AND RYAN SOLAVA

**ABSTRACT.** Chen, Kitaev, Mütze, and Sun introduced universal partial words, a generalization of universal words. Universal partial words allow for a wild-card character  $\diamond$  which can cover any letter in the alphabet. We address some conjectures posed in the introductory paper. In addition, we give structural results for non-binary alphabets, as well as provide an explicit construction for a family of universal partial words for alphabets of even size.

## 1. INTRODUCTION

*Universal cycles* of a wide variety of combinatorial structures have been well studied [6]. The best known examples are the *de Bruijn cycles*, cyclic sequences over an alphabet  $A$  which contain each word of length  $n$  as a substring exactly once. These arise from finding Hamilton cycles in the *de Bruijn graph*.

We denote the set of all words of length  $n$  over an alphabet  $A$  by  $A^n$ . A *universal word* for  $A^n$  is a word  $w$  such that each word in  $A^n$  appears exactly once as a consecutive substring of  $w$ . For example, 0001011100 is a universal word for  $\{0,1\}^3$ . The length of a universal word for  $A^n$  is  $|A|^n + n - 1$ . It is known [5] that universal words for  $A^n$  exist for any  $n$ , but a brute-force search for universal words would quickly become intractable as it would require checking  $|A|^{|A|^n + n - 1}$  words.

Partial words are sequences of symbols from  $A \cup \{\diamond\}$ , where  $\diamond \notin A$  can correspond to any letter of  $A$ . Partial words are natural objects in coding theory and theoretical computer science. As an example, when representing DNA and RNA in computing as a string, the  $\diamond$  character can take the place of any unknown nucleotide. There are also applications to molecular biology and data communication [2].

In 2016, Chen, Kitaev, and Sun [5] introduced the universal partial word for  $A^n$ , generalizing the universal word. For partial words  $u$  and  $v$ , we say that  $u \subset v$  (or  $u$  is a factor of  $v$ , or  $v$  covers  $u$ ), if there exists  $i$  such that  $u_j = v_{i+j}$  for  $1 \leq j \leq |u|$  whenever  $v_{i+j} \in A$ .

**Definition 1.1.** A *universal partial word* for  $A^n$  is a partial word  $w$  that covers each word in  $A^n$  exactly once.

For example,  $\diamond\diamond 0111$  is a universal partial word for  $\{0,1\}^3$ . Universal partial words may be useful in questions related to storing information in compact form, because allowing  $\diamond$ 's decreases the length required to cover all words of length  $n$ .

---

Research partially supported by NSF-DMS grant #1604458, "Rocky Mountain - Great Plains Graduate Research Workshops in Combinatorics".

In [5], Chen et al. investigated the existence and nonexistence of universal partial words over a binary alphabet containing exactly one or two consecutive  $\diamond$ 's. We present some preliminaries in Section 2. In Section 3, we resolve two conjectures from [5] and develop some machinery that helps motivate the consideration of such words over non-binary alphabets. In Section 4, we prove the following structural result about universal partial words over non-binary alphabets.

**Theorem (4.1).** *Let  $w$  be a universal partial word for  $A^n$  with  $|A| \geq 3$ . If  $w_i = \diamond$ , then  $w_j = \diamond$  for all  $i \equiv j \pmod n$ .*

As a consequence of the above theorem, the diamond structure is much more constrained than in the binary case, which both restricts the possible lengths and gives rise to number theoretic conditions limiting their existence. In addition, we show that non-binary alphabets impose a pseudocyclic structure on universal partial words. In Section 5, we show that there are no non-trivial universal partial words for  $n \leq 3$  over non-binary alphabets, but give an explicit construction for  $n = 4$  and any even alphabet size  $a$ . Finally, in Section 6, we discuss some open questions.

## 2. PRELIMINARIES

An *alphabet*  $A$  is a set of symbols, which we call *letters*. Throughout this paper, we will denote the size of the alphabet  $A$  by  $a$  and assume without loss of generality that any alphabet of size  $a$  is  $\{0, 1, \dots, a - 1\}$ .

A *word* over an alphabet  $A$  is a sequence of letters. A *partial word* is a sequence of symbols from  $A \cup \{\diamond\}$ . Note that a word cannot contain a  $\diamond$ ; we sometimes use the term *total word* instead of *word* for emphasis.

Trivially,  $\diamond^n$  is a universal partial word for  $A^n$  for any  $A$  and  $n$ . A universal partial word  $w$  is called *trivial* if all of its characters are diamonds ( $w = \diamond^n$ ) or none are ( $w$  is a universal word for  $A^n$ ). We are interested in the existence of nontrivial universal partial words.

Given an universal partial word  $w$ , a *window* of  $w$  is a string of  $n$  consecutive characters in  $w$ , and a *pane* is a string of  $n - 1$  consecutive characters in  $w$ . The *diamondicity* of  $w$  is the number of diamonds that appear in any window of  $w$ . (Note that diamondicity is well-defined when  $|A| \geq 3$  by Theorem 4.1.) The *frame* of a partial word is the total word over  $\{_, \diamond\}$  obtained by replacing all letters from  $A$  by the “\_” character. A *window frame* is the frame of a window.

Borders and periods are related and fundamental concepts in the study of combinatorics on words [3]. A partial word  $w$  has a *border*  $x$  of length  $k$  if both the first  $k$  symbols and the last  $k$  symbols of  $w$  cover  $x$ . A *period* of a word  $w$  is a positive integer  $p$  such that  $w_i = w_j$  whenever  $i \equiv j \pmod p$ . Borders and periods have the following well known relationship.

**Theorem 2.1** (Folklore, e.g. [1]). *A word  $w$  has period  $p \in [|w| - 1]$  if and only if it has a border of length  $|w| - p$ .*

### 3. EXISTENCE AND NONEXISTENCE OF UNIVERSAL PARTIAL WORDS OVER A BINARY ALPHABET

In [5], Chen et. al. initially considered universal partial words over binary alphabets. The authors presented two conjectures in their original paper. The first of these we answer in the negative and provide their new modified conjecture. The second of these we prove later in this section.

**Conjecture 3.1** (Conjecture 1 from [5]). *Let  $n \geq 2$ . An universal partial word for  $\{0, 1\}^n$  of the form  $u \diamond v$  (with  $u$  and  $v$  words) exists if and only if either  $|u| \leq n - 2$  or  $|v| \leq n - 2$ .*

We have produced counterexamples to this conjecture, one of which is the following.

**Example 3.2.**  $w = 00110 \diamond 001011110$  is a universal partial word for  $\{0, 1\}^4$ .

After corresponding our counterexample to the authors of [5], they amended their conjecture to the following.

**Conjecture 3.3** (Conjecture 8 from [4]). *There exists an upword for  $\{0, 1\}^n$  with a single diamond in position  $k$ , where  $k \neq n$ , outside of a small set of counterexamples for small  $n$ .*

Chen et. al. also made the following conjecture in their original paper, [5].

**Conjecture 3.4** (Conjecture 2 from [5]). *For  $n \geq 4$ , there exists no universal partial word of the form  $u \diamond \diamond v$  for any binary words  $u$  and  $v$ .*

We will use the remainder of this section to prove Conjecture 3.4. We will also extend Conjecture 3.4 both by considering words with more than two diamonds and by considering words over alphabets of larger size.

We use the following lemma, which restricts the periodicity of any  $n - k$  letters following  $k$  consecutive diamonds in a universal partial word.

**Lemma 3.5.** *There does not exist a universal partial word  $w$  containing the substring  $u \diamond^k v$  where  $u$  and  $v$  are words,  $k \geq 2$ ,  $|v| = n - k$ ,  $v$  has period  $p \leq k$ , and  $|u| = p$ .*

*Proof.* Suppose such a  $w$  exists, and let  $v = v_1 v_2 \cdots v_{n-k}$ , with  $v_i \in A$  for all  $i \in [n - k]$ . Since  $v$  has period  $p$ , by Theorem 2.1,  $v_1 v_2 \cdots v_{n-k-p} = v_{p+1} v_{p+2} \cdots v_{n-k}$ , and so  $\diamond^p v_1 v_2 \cdots v_{n-k-p}$  covers  $v$ . Thus, the string  $u 0^{k-p} v$  is covered by both the window starting with  $u$  ( $u \diamond^k v_1 v_2 \cdots v_{n-k-p}$ ) and the window beginning at the first diamond ( $\diamond^k v$ ). Therefore,  $w$  is not a universal partial word.  $\square$

Lemma 3.5 and the following result hold for any alphabet size. However, a stronger result for non-binary alphabets will be stated in Section 4, so we address only the binary case in the following theorem.

**Theorem 3.6.** *Over binary alphabets, there does not exist a universal partial word  $w = u \diamond^k v$  where  $u$  and  $v$  are nonempty words over  $A$ ,  $k \geq 2$ , and  $|v| \geq n$ .*

*Proof.* Suppose such a  $w$  exists, and let  $w = w_1 w_2 \cdots w_m$  and  $v = v_1 v_2 \cdots v_\ell$ .

There are  $2^k$  words of length  $n$  that begin  $v_1v_2 \cdots v_{n-k}$ , all of which must be covered by  $w$ . One of these words is  $v_1v_2 \cdots v_{n-k} \cdots v_n$ .

Since  $\diamond^k v$  covers all words of length  $n$  ending in  $v_1v_2 \cdots v_{n-k}$ ,  $v_1v_2 \cdots v_{n-k}$  is not covered elsewhere in  $w$  with  $k$  characters preceding it, so the only other places it can begin in  $w$  are the first  $k$  positions. Thus, the remaining  $2^k - 1$  words beginning with  $v_1v_2 \cdots v_{n-k}$  must be covered by windows that begin in the first  $k$  positions of  $w$ .

For each such position  $i \in [k]$ , let  $V_i$  be the number of words in  $A^n$  beginning with  $v_1v_2 \cdots v_{n-k}$  that are covered by  $W_i := w_iw_{i+1} \cdots w_{i+n-1}$ , the window beginning at position  $i$  of  $w$ .

Then  $\sum_{i=1}^k V_i = 2^k - 1$ , and for all  $i$ ,  $V_i$  is either 0 or  $2^j$  if the first  $n-k$  symbols of  $W_i$  cover  $v_1v_2 \cdots v_{n-k}$ , where  $j$  is the number of  $\diamond$ 's in the last  $k$  characters of  $W_i$ . Since  $V_i \leq \sum_{i=1}^k V_i = 2^k - 1$ ,  $j < k$ .

Now,  $2^k - 1$  can be uniquely written as a sum of powers of 2 as  $\sum_{i=0}^{k-1} 2^i$ , so we must have that  $V_1, \dots, V_k$  are  $2^0, 2^1, \dots, 2^{k-1}$  in some order. Since  $V_{i+1} \neq V_i$  and  $W_i$  has at most either one more or one less  $\diamond$  than  $W_{i+1}$ ,  $V_{i+1} = 2V_i$  or  $\frac{1}{2}V_i$  for all  $i \in [k-1]$ . Also, since we use each power exactly once, if  $V_{i+1} = 2V_i$ , then  $V_{i+2} = 2V_{i+1}$ . Thus either, for all  $i$ ,  $V_i = 2^{i-1}$  and  $w = v_1v_2 \cdots v_{n-k}w_{n-k+1}w_{n-k+2} \cdots w_n \diamond^k v$ , or  $V_i = 2^{k-i}$  and  $w = v_1v_2 \cdots v_{n-k-1} \diamond^k v$ .  
Case 1:  $w = v_1v_2 \cdots v_{n-k}w_{n-k+1}w_{n-k+2} \cdots w_n \diamond^k v$

Since  $V_i > 0$  for all  $i$ ,  $v_1 = v_2 = \cdots = v_{n-k}$ . Thus,  $v_1v_2 \cdots v_{n-k}$  has period 1. Since  $|u| \geq 1$ , by Lemma 3.5,  $w$  is not a universal partial word.

Case 2:  $w = v_1v_2 \cdots v_{n-k-1} \diamond^k v$

Since,  $|u| \geq 1$ , this case is only possible if  $n - k - 1 \geq 1$ . Since  $V_i > 0$  for all  $i$ ,  $v_1 = v_2 = \cdots = v_{n-k-1}$ . If  $n - k - 1 \geq 2$ , then  $W_1$  and  $W_2$  both cover  $v_1^{n-k-1}v_{n-k}0^{k-2}v_1^2 = v_1v_2 \cdots v_{n-k}0^{k-2}v_1v_2$ .

Thus,  $n - k - 1 = 1$ , and  $w = v_1 \diamond^k v = v_1 \diamond^{n-2} v$ . Without loss of generality, assume  $v_1 = 0$ . Then,  $w = 0 \diamond^{n-2} 0v_2v_3 \cdots v_\ell$ . Now,  $W_1$  covers all words beginning and ending with 0. Thus,  $v_n = 1$ , and every letter  $n-1$  positions after a diamond is a 1, so  $w = 0 \diamond^{n-2} 01^{n-1}v_{n+1} \cdots v_\ell$ . Next, since  $w$  covers  $1^n$ ,  $v_{n+1} = 1$ . Otherwise,  $01^{n-1}$  would be covered twice in  $w$ . Thus,  $w = 0 \diamond^{n-2} 01^n v_{n+2} \cdots v_\ell$ . Since  $\diamond^{n-2} 01$  is covered by  $W_2$ ,  $01$  cannot appear elsewhere in  $w$  with  $k$  characters preceding it. Thus,  $v_{n+2} = v_{n+3} = \cdots = v_\ell = 0$ . Since  $0^n$  is covered by  $W_1$ ,  $w = 0 \diamond^{n-2} 01^n 0^p$  where  $p \leq n-1$ . Now, if  $n$  is even,  $(10)^{\frac{n}{2}}$  is not covered by  $w$ , and if  $n$  is odd,  $(10)^{\frac{n-1}{2}}0$  is not covered by  $w$ . Thus,  $w$  is not a universal partial word.  $\square$

Next, we state and prove two lemmas in order to prove Theorem 3.9, which is a generalization of Conjecture 3.4 that has been extended to include larger alphabets.

**Lemma 3.7.** *For  $a \geq 2$ , there does not exist a universal partial word  $w = u \diamond^{n-k} v$  where  $k \leq \min\{n/2, |u|, |v|\}$  and  $u$  and  $v$  are words.*

*Proof.* Let  $w = u \diamond^{n-k} v$  where  $k \leq \min\{n/2, |u|, |v|\}$ . Let  $u = u' u_1 \cdots u_k$  and  $v = v_1 \cdots v_k v'$ . Then the word  $u_1 \cdots u_k 0^{n-2k} v_1 \cdots v_k$  is covered twice.  $\square$

**Lemma 3.8.** *Let  $w = \diamond \diamond v$ , where  $v$  is a word. If  $n \geq 4$ ,  $|v| \geq n - 2$  and  $a \geq 2$ , then  $w$  is not a universal partial word.*

*Proof.* Let  $v' = v_1 \dots v_{n-2}$ . All words ending with  $v'$  are covered by  $\diamond \diamond v'$ , so  $v'$  cannot appear later in  $v$ . Thus the  $a^2$  words beginning with  $v'$  must appear starting in positions 1, 2, or 3 of  $w$ . Since  $n - 2 \geq 2$ ,  $w$  can contain at most 3 of these words, so  $w$  is not a universal partial word.  $\square$

Together, these results lead us to the proof of a stronger version of Conjecture 3.4:

**Theorem 3.9.** *There does not exist a universal partial word  $w = u \diamond \diamond v$  over any alphabet with  $a \geq 2$ , when  $n \geq 4$  and  $u$  and  $v$  are (possibly empty) words.*

*Proof.* We proceed by contradiction. Assume  $w = u \diamond \diamond v$  is a universal partial word.

By Lemmas 3.7 and 3.8 and the reversal property,  $|u|, |v| \leq n - 1$ . There are at most  $a$  words covered by  $u$  and the first  $\diamond$ . There are at most  $(n - 1)a^2$  words covered using two  $\diamond$ s (starting at each position of  $u$  except the first and starting at the first  $\diamond$ ). Finally, there are at most  $a$  words covered by the second diamond and  $v$ . Since  $|u|, |v| \leq n - 1$ , no string can end before or begin after the  $\diamond$ s. Thus, there are at most  $2a + (n - 1)a^2$  words covered by  $w$ .

Now,  $2a + (n - 1)a^2 < a + na^2$  since  $a^2 > a$  and  $a \geq 2$ . If  $a \geq 3$  and  $n \geq 4$ , then  $w$  does not cover all of the words in  $A^n$ . Similarly, if  $a = 2$  and  $n \geq 5$ ,  $w$  again does not cover enough words.

Now, for  $a = 2, n = 4$ , consider  $w$  with  $|u|, |v| = n - 1 = 3$ , so  $w = rst \diamond \diamond xyz$ . Without loss of generality, let  $z = 0$ . By Lemma 3.5,  $y \neq 0$ , so  $y = 1$ . Similarly if  $t = 0$ , then  $s = 1$ , and then 1001 occurs starting at both  $s$  and the first  $\diamond$ . Thus,  $t = 1$ . By periodicity,  $s = 0$ , but then 0101 appears in the word beginning at both  $s$  and the first  $\diamond$ . Thus, such a  $w$  does not exist.  $\square$

In an updated version, Chen et. al. provide an independent proof resolving their conjecture.

#### 4. STRUCTURAL CONDITIONS OVER NON-BINARY ALPHABETS

Many of the above results hold for all alphabet sizes. If we restrict our attention to the non-binary case, we are able to produce structural results which give more succinct proofs for the non-binary cases in Section 3.

**Theorem 4.1.** *Let  $w$  be a universal partial word for  $A^n$  with  $a \geq 3$ . If  $w_i = \diamond$ , then  $w_j = \diamond$  for all  $i \equiv j \pmod n$ . In other words, the frame of any universal partial word has period  $n$ .*

*Proof.* We will show that if  $w_i = \diamond$ , then  $w_{i+n} = \diamond$  for any  $i \in [|w| - n]$ . Since the reverse of  $w$  must also be a universal partial word, this is sufficient to obtain the theorem.

Suppose  $w_i = \diamond$  and  $w_{i+n} \in A$  for the sake of contradiction. Without loss of generality, let  $w_{i+n} = 0$ . Let  $v$  be any total word covered by  $w_{i+1}w_{i+2} \dots w_{i+n-1}$ ; note  $w$  has a substring that covers  $\diamond v 0$  and ends in 0, and the word  $v 0$  has length  $n$ . Since  $w$  is a universal partial word, it must cover the words  $v 1$  and  $v 2$  exactly once. Either the words  $v 1$  and  $v 2$  are both covered by  $w_1 \dots w_n$  or not.

In the first case,  $w_n = \diamond$ , so  $w_1 \dots w_n$  also covers the word  $v 0$ . This instance of the word  $v 0$  is different from the one that appears following  $w_i$  because it appears at the beginning

of  $w$ . Therefore the word  $v0$  is covered twice in  $w$ , contradicting that  $w$  is a universal partial word.

In the second case, without loss of generality, the word  $v1$  is covered by  $w$  but not covered by  $w_1 \cdots w_n$ . Since this instance of the word  $v1$  is not at the beginning of  $w$ , let  $x \in A \cup \{\diamond\}$  be the character of  $w$  immediately preceding it. Then the word  $xv$  is covered twice: once in this substring of  $w$ , which covers  $xv1$ , and once in the aforementioned substring of  $w$  that covers  $\diamond v0$  and ends in 0. This contradicts that  $w$  is a universal partial word.  $\square$

As a corollary, we obtain the following non-binary analogue of Lemma 3.6.

**Corollary 4.2.** *There does not exist a universal partial word  $w = u \diamond^k v$  where  $u$  and  $v$  are words,  $k \geq 2$ ,  $|v| \geq n$ , and  $a \geq 3$ .*

Theorem 4.1 makes our definition of diamondicity well-defined.

**Definition 4.3.** For  $w$  a universal partial word for  $A^n$ , with  $a \geq 3$ , the diamondicity of  $w$  is the number of diamonds in each window. The frames of  $w$  are a repeating pattern of cyclic shifts of the first frame.

For binary words, knowing the number of diamonds is not enough to determine the length of the word. However, using diamondicity, we can find the length of universal partial words over larger alphabets.

**Corollary 4.4.** *If  $w$  is a universal partial word for  $A^n$  with diamondicity  $d$ , then  $|w| = a^{n-d} + n - 1$ .*

*Proof.* Each window contains  $d$  diamonds, by Theorem 4.1. Therefore, each window covers  $a^d$  words. Since  $w$  is a universal partial word, it must cover all  $a^n$  words, so  $w$  contains  $\frac{a^n}{a^d} = a^{n-d}$  windows. The last window contains  $n - 1$  characters that do not themselves start windows, so  $|w| = a^{n-d} + n - 1$ .  $\square$

Words over binary alphabets containing a single diamond were studied in [5]. In fact, over non-binary alphabets, such words do not exist.

**Proposition 4.5.** *For  $a \geq 3$ , there does not exist a universal partial word  $w$  containing exactly one diamond.*

*Proof.* Assume that  $n \geq 4$ ,  $a \geq 3$ , and  $w = u \diamond v$  where  $u$  and  $v$  are words. By diamondicity, we know that  $|u|, |v| \leq n - 1$ . Also by diamondicity, we know that  $|w| = a^{n-1} + n - 1$ . Therefore

$$a^{n-1} + n - 1 = |u| + |v| + 1 \leq n - 1 + n - 1 + 1 = 2n - 1$$

and therefore  $a^{n-1} \leq n$ . This is a contradiction. The case where  $n < 4$  is shown in Proposition 5.1.  $\square$

While we only consider the one diamond case, Theorem 4.1 shows that the correct thing to consider is not the number of diamonds, but rather the density of diamonds, since there are roughly  $(d/n) \cdot a^{n-d}$  diamonds in a universal partial word over a non-binary alphabet.

Classical universal words are cyclic but for universal partial words over binary alphabets, this need not be the case. For non-binary alphabets, we can achieve a similar notion of cyclicity to the classical case.

**Proposition 4.6.** *If  $w$  is a universal partial word for  $A^n$  with  $a \geq 3$ , then the first  $n - 1$  characters of  $w$  equal the last  $n - 1$  characters of  $w$ . (In other words, as a (non-partial) word over the extended alphabet  $A \cup \{\diamond\}$ ,  $w$  has a border of length  $n - 1$ .)*

*Proof.* Let  $w$  be a universal partial word for  $A^n$  with  $a \geq 3$ . Let  $v$  be a word of length  $n - 1$  covered by  $w_1 \cdots w_{n-1}$ . We show in two cases that the last  $n - 1$  characters of  $w$  cover  $v$ .

First, if  $w_n = \diamond$ , all  $a$  words beginning with  $v$  are covered by  $w_1 \cdots w_n$ . If  $v$  was covered elsewhere in  $w$ , except for the end of  $w$ , then that string and the character immediately following it would cover a word beginning with  $v$  also covered by  $w_1 \cdots w_n$ . Therefore, for  $w$  to cover the  $a$  words ending with  $v$ , the last  $n - 1$  character of  $w$  cover  $v$ .

On the other hand, if  $w_n \neq \diamond$ , exactly one word beginning with  $v$  is covered by  $w_1 \cdots w_n$ . The remaining  $a - 1$  words beginning with  $v$  must be covered by strings of  $w$  that are not immediately followed by  $\diamond$ , as this would duplicate  $v_1 v_2 \cdots v_{n-1} w_n$ . These words are covered by exactly  $a - 1$  other strings  $w_{1_1} \cdots w_{1_{n-1}}, \dots, w_{(a-1)_1} \cdots w_{(a-1)_{n-1}}$  in  $w$  that cover  $v$  and are followed by a letter in  $A$ . By Proposition 4.1, these strings cannot be preceded by  $\diamond$ , and thus  $w_{1_1-1} w_{1_1} \cdots w_{1_{n-1}}, \dots, w_{(a-1)_1-1} w_{(a-1)_1} \cdots w_{(a-1)_{n-1}}$  cover  $a - 1$  distinct words ending with  $v$ . The remaining word ending with  $v$  must be covered by the last  $n$  characters of  $w$ , so the last  $n - 1$  characters of  $w$  cover  $v$ .

Thus the last  $n - 1$  characters of  $w$  cover all words covered by the first  $n - 1$  characters, and since the reversal of  $w$  is also a universal partial word for  $A^n$ , the first  $n - 1$  characters cover all words covered by the last  $n - 1$  characters. Hence the first  $n - 1$  characters equal the last  $n - 1$  characters.  $\square$

By Corollary 4.4, the higher the diamondicity of a word, the shorter it is. However, as diamondicity increases it becomes harder to avoid covering words multiple times. In fact, it is possible to bound the potential diamondicities of a universal partial word in terms of  $n$  as shown in the following proposition.

**Proposition 4.7.** *For every  $k$ , if  $n \geq k(k - 1) + 1$ , it is impossible to construct a word with diamondicity  $d \geq n - k$  over  $A^n$ .*

*Proof.* Suppose  $w$  is a universal partial word with diamondicity  $n - k$ . Consider the first frame  $f$  of  $w$ . Let  $x_1, x_2, \dots, x_k$  be the positions of the  $\_$ 's in  $f$ . If there exists another frame  $f'$  in  $w$  which shares no  $\_$ 's with  $f$  (i.e. the positions where  $f$  has  $\_$ 's,  $f'$  has diamonds, and vice versa), then they clearly cover a word twice.

Suppose the  $i$ th frame does not have this property. Then there is a  $j$  so that  $x_i + k = x_j$ . Note that when we view this cyclically  $x_j + (n - k) = x_i$ . There are at most  $k(k - 1)$  distinct distances between  $x_i$ 's and  $x_j$ 's, since each pair has a distance  $d$  and  $n - d$  between them. Call this set of distances  $\mathcal{K}$ .

So if  $n \geq k(k - 1) + 1$ , by the pigeonhole principle there is at least one  $d < n$  with  $d \notin \mathcal{K}$ . Thus the  $d$ th frame and  $f$  share no  $\_$ 's, so they cover a word twice. This is a contradiction

to  $w$  being a universal partial word, so  $n < k(k-1) + 1$  in order for  $n-k$  diamondicity to be possible.  $\square$

This proof method is insufficient for giving a fractional bound on diamondicity, since it is possible to construct frames for a given density for large enough  $n$  which do not clearly cover any strings twice.

We would like to be able to show the nonexistence of universal partial words based only on the parameters  $a, n$ , and  $d$ . We will take advantage of the cyclic nature of the frames.

**Lemma 4.8.** *Cyclically shifting a word  $f$   $i$  times yields  $f$  if and only if there is a word  $g'$  such that  $f = (g')^s$  for some  $s \in \mathbb{N}$ , where  $|g'| = i$ .*

*Proof.* If  $f = (g')^s$  for some  $s \in \mathbb{N}$ , where  $|g'| = i$ , then cyclically shifting  $f = (g')(g')^{s-1}$   $i$  times yields  $(g')^{s-1}(g') = g^s = f$ . For the reverse direction, suppose cyclically shifting  $f$   $i$  times yields  $f$ . Then  $f_1 f_2 \cdots f_n = f_{i+1} f_{i+2} \cdots f_n f_1 f_2 \cdots f_i$ . Let  $g' = f_1 f_2 \cdots f_i$ . Note  $f$  has period  $i$  (since  $f_j = f_{i+j \bmod n}$  for all  $j \in [n]$ ) and begins and ends with  $g'$ , so  $i|n$  and  $f = (g')^{n/i}$ , where  $n/i \in \mathbb{N}$ .  $\square$

Using Theorem 2.1, we can prove the following theorem.

**Theorem 4.9.** *If  $w$  is a universal partial word for  $A^n$  with  $a \geq 3$ ,  $f$  is the frame of the first window of  $w$ , and  $f'$  is the shortest frame such that  $f = (f')^t$  for some  $t \in \mathbb{N}$ , then  $|f'| \mid \gcd(a^{n-d}, n)$ .*

*Proof.* First, we note that  $t|f'| = |f| = n$ , so  $|f'| \mid n$ .

Next, we will show that  $|f'| \mid a^{n-d}$ .

The length of  $w$  is  $N = a^{n-d} + n - 1$ . By Proposition 4.6 and Theorem 2.1,  $w$  considered as a word over  $A \cup \{\diamond\}$  has period  $N - (n-1) = a^{n-d}$ . In particular, the frame of  $w$  has period  $a^{n-d}$ , so cyclically shifting the frame of the first window of  $w$   $a^{n-d}$  times yields the same frame.

As consequences of Lemma 4.8 and the definition of  $|f'|$ , we can conclude that (a) if we shift  $f$   $|f'|$  times, we get the same frame, and that (b) if we shift  $f$  fewer than  $|f'|$  times, we cannot get the same frame.

Let  $r = a^{n-d} \bmod |f'|$ . This remainder  $r$  is less than  $|f'|$ , and shifting  $f$   $r$  times must yield the same frame because by Lemma 4.8 it is equivalent to shifting  $f$   $a^{n-d}$  times, which as shown above yields the same frame. Therefore  $r = 0$ , i.e.  $|f'| \mid a^{n-d}$ .  $\square$

This gives rise to some immediate number-theoretic corollaries which allow us to eliminate many combinations of  $a, n$ , and  $d$ .

**Corollary 4.10.** *For  $a \geq 3$ , if  $\gcd(a, n) = 1$ , then there are no non-trivial universal partial words for  $A^n$ .*

*Proof.* Since  $\gcd(a, n) = 1$ ,  $\gcd(a^{n-d}, n) = 1$ . Then, using the notation of Theorem 4.9,  $|f'| = 1$ , so  $f = \diamond^n$  or  $f = \_n$ . Thus,  $w$  is trivial.  $\square$



Furthermore, it is easy to show when  $\gcd(a^{n-d}, n) = 2$ , that no non-trivial upwords exist.

**Corollary 4.11.** *If  $\gcd(n, d) = 1$ , then  $n|a^{n-d}$ . In particular, if  $d = 1$ , then  $n|a^{n-1}$ .*

*Proof.* By Theorem 4.9,  $n'|a^{n-d}$ , where  $n' = |f'|$ . Let  $d'$  be the number of diamonds in  $f'$ . We have that  $n't = n$  and  $d't = d$ , so if  $\gcd(n, d) = 1$ , then  $t = 1$ . Thus  $n' = n$ , and  $n|a^{n-d}$  by Theorem 4.9.  $\square$

**Corollary 4.12.** *If  $\gcd(a^{n-d}, n) = p$  for some prime  $p$ , then  $d \in \{kn/p : k \in [p-1]\}$ .*

*Proof.* By Theorem 4.9,  $|f'| = 1$  or  $p$ ; assuming  $w$  is non-trivial, we have  $|f'| = p$ . Then,  $n = pt$ . Let  $d'$  be the number of diamonds in  $f'$ , so  $d = d't = d'n/p$ . Note  $d' \in [p-1]$  as  $f'$  must have at least one letter and at least one diamond for  $w$  to be non-trivial.  $\square$

## 5. CONSTRUCTION OF A UNIVERSAL PARTIAL WORD

Given the results of Section 4, it is tempting to believe that universal partial words do not exist for any non-binary alphabets. This is in fact the case for small  $n$ .

**Proposition 5.1.** *For  $a \geq 3$ , there does not exist a non-trivial universal partial word  $w$  for  $n \leq 3$ .*

*Proof.* It is clear that there is no nontrivial universal partial word for  $n = 1$ .

For  $n = 2$ , assume  $w$  is non-trivial. By Theorem 4.1,  $d = 1$ , and by Corollary 4.4,  $|w| = a + 1 \geq 4$ . Now,  $w = \diamond x_2 \diamond x_4 \diamond \dots$  or  $w = x_1 \diamond x_3 \diamond x_5 \diamond \dots$ . In the first case, the first three characters cover  $x_2 x_2$  twice. In the second case, characters two through four cover  $x_3 x_3$  twice. Therefore,  $w$  is not a universal partial word.

For  $n = 3$ , assume  $w$  is non-trivial. By Theorem 4.1,  $w$  must either contain the frame  $\diamond \_ \_$  or  $\diamond \diamond \_$ .

In the first case, consider the string 000. Our word  $w$  can't contain the string  $\diamond 00 \diamond$ , since that would cover 000 twice, so  $w$  must contain the string  $\diamond x0 \diamond 0y \diamond$  to cover 000. But this string covers  $x0y$  twice. Therefore,  $w$  is not a partial word.

In the second case,  $|w| = a + 2 \geq 5$  and must contain either  $\diamond \diamond x \diamond \diamond$ ,  $x \diamond \diamond y \diamond$ , or  $\diamond x \diamond \diamond y$ . The first four characters of each of these partial words cover either  $xxx$  or  $xyx$  twice. Therefore,  $w$  is not a universal partial word.  $\square$

While these small  $n$  are not fruitful, not only are we able to find non-trivial examples for  $n = 4$ , we can construct a family of universal partial words for any alphabet size of even size. Note that by Corollary 4.10, there are no non-trivial universal partial words for  $n = 4$  when  $a$  is odd, for then  $\gcd(a, 4) = 1$ .

**Theorem 5.2.** *Let  $A = (0, 1, \dots, a-1)$ , where  $a$  is even.*

(1) *Construct the following sequence of  $a^3/4$  symbols:*

$$\underbrace{0, 1, 0, 1, \dots, 0, 1}_{a^2/2 \text{ symbols}}, \underbrace{2, 3, 2, 3, \dots, 2, 3}_{a^2/2 \text{ symbols}}, \dots, \underbrace{a-2, a-1, a-2, a-1, \dots, a-2, a-1}_{a^2/2 \text{ symbols}}.$$

*Call this sequence  $\langle x_i \rangle$ , where  $0 \leq i < a^3/4$ .*

(2) Construct the following sequence of  $a^2/2$  symbols:

$$\underbrace{0, 1, 0, 1, \dots, 0, 1}_a \underbrace{2, 3, 2, 3, \dots, 2, 3}_a \dots \underbrace{a-2, a-1, a-2, a-1, \dots, a-2, a-1}_a.$$

Then repeat that sequence  $a/2$  times to get a sequence of  $a^3/4$  symbols, and call the resulting sequence  $\langle y_i \rangle$ , where  $0 \leq i < a^3/4$ .

(3) Construct the following sequence of  $a$  symbols:

$$1, 0, 3, 2, 5, 4, \dots, a-1, a-2.$$

Then repeat that sequence  $a^2/4$  times to get a sequence of  $a^3/4$  symbols, and call the resulting sequence  $\langle z_i \rangle$ , where  $0 \leq i < a^3/4$ .

(4) For  $0 \leq i < a^3/4$ , let  $w_i$  be the word  $x_i y_i z_i$ .

(5) Take  $u = w_0 \diamond w_1 \diamond w_2 \diamond \dots \diamond w_{(a^3/4)-1} \diamond w_0$ .

Then  $u$  is a universal partial word for  $A^4$ .

*Proof.* Each  $w_i$  has length 3 and there are  $a^3/4$  diamonds, so

$$|u| = 3((a^3/4) + 1) + a^3/4 = a^3 + 3.$$

This is the length of a universal partial word with  $n = 4$  and diamondicity 1, so it is sufficient to show that no word is covered twice.

Suppose  $u$  covers  $v_1 v_2 v_3 v_4$  twice. Let  $v$  and  $v'$  be the two windows of  $u$  which cover  $v_1 v_2 v_3 v_4$ . Either  $v$  and  $v'$  have a diamond in the same position or in a different position.

Case 1: Suppose  $v$  and  $v'$  have the same frame. Let us consider when  $v = x_i y_i z_i \diamond$  and  $v' = x_j y_j z_j \diamond$  with  $i < j$ . Suppose  $x_i = x_j = c$ . Then we must have that

$$\frac{\lfloor c/2 \rfloor a^2}{2} \leq i, j \leq \frac{(\lfloor c/2 \rfloor + 1) a^2}{2}.$$

Within this range,  $y_i = y_j$  implies that  $j - i < a$ , but  $z_i = z_j$  implies that  $j - i \geq a$ . This is a contradiction, so  $v$  and  $v'$  do not cover the same word.

Note that if the diamond were further to the left in  $v$  and  $v'$ , not all of the indices would be the same (e.g.  $z_i \diamond x_{i+1} y_{i+1}$ ), but this would increment both  $i$  and  $j$ , so their difference would still be the same.

Case 2: Suppose  $v$  and  $v'$  have diamonds in different positions. In general, we have that  $i \equiv x_i \equiv y_i \pmod{2}$  and  $i \not\equiv z_i \pmod{2}$ .

- If  $v = x_i y_i z_i \diamond$  and  $v' = y_j z_j \diamond x_{j+1}$ , then  $y_j = x_i$  and  $y_i = z_j$ , so

$$y_j \equiv x_i \equiv y_i \equiv z_j \pmod{2},$$

a contradiction.

- If  $v = x_i y_i z_i \diamond$  and  $v' = z_j \diamond x_{j+1} y_{j+1}$ , then  $x_i = z_j$  and  $z_i = x_{j+1}$ , so

$$z_i \equiv x_{j+1} \equiv z_j \equiv x_i \pmod{2},$$

a contradiction.

- If  $v = x_i y_i z_i \diamond$  and  $v' = \diamond x_j y_j z_j$ , then  $z_i = y_j$  and  $y_i = x_j$ , so

$$z_i \equiv y_j \equiv x_j \equiv y_i \pmod{2},$$

a contradiction.

- If  $v = y_i z_i \diamond x_{i+1}$  and  $v' = z_j \diamond x_{j+1} y_{j+1}$ , then  $y_i = z_j$  and  $y_{j+1} = x_{i+1}$ , so

$$y_i \equiv z_j \equiv y_{j+1} \equiv x_{i+1} \pmod{2},$$

a contradiction.

- If  $v = y_i z_i \diamond x_{i+1}$  and  $v' = \diamond x_j y_j z_j$ , then  $z_i = x_j$  and  $x_{i+1} = z_j$ , so

$$x_j \equiv z_i \equiv x_{i+1} \equiv z_j \pmod{2},$$

a contradiction.

- If  $v = z_i \diamond x_{i+1} y_{i+1}$  and  $v' = \diamond x_j y_j z_j$ ,  $x_{i+1} = y_j$  and  $y_{i+1} = z_j$ , so

$$z_j \equiv y_{i+1} \equiv x_{i+1} \equiv x_j \pmod{2},$$

a contradiction.

Note that since the fact that  $i < j$  is never used, the arguments are reflexive, so these are all of the cases.

So  $v_1 v_2 v_3 v_4$  is not covered twice by  $u$ , therefore  $u$  is a universal partial word.  $\square$

We give an example of this construction for  $a = 4$ .

**Example 5.3.** The string

$001 \diamond 110 \diamond 003 \diamond 112 \diamond 021 \diamond 130 \diamond 023 \diamond 132 \diamond 201 \diamond 310 \diamond 203 \diamond 312 \diamond 221 \diamond 330 \diamond 223 \diamond 332 \diamond 001$

is a universal partial word for  $\{0, 1, 2, 3\}^4$ .

Here

- $\langle x_i \rangle = \langle 0, 1, 0, 1, 0, 1, 0, 1, 2, 3, 2, 3, 2, 3, 2, 3 \rangle$
- $\langle y_i \rangle = \langle 0, 1, 0, 1, 2, 3, 2, 3, 0, 1, 0, 1, 2, 3, 2, 3 \rangle$
- $\langle z_i \rangle = \langle 1, 0, 3, 2, 1, 0, 3, 2, 1, 0, 3, 2, 1, 0, 3, 2 \rangle$ .

## 6. OPEN PROBLEMS

While we have constructed an infinite family of universal partial words over  $A^4$ , increasing  $n$  to 5 already makes brute force searches infeasible, even when accounting for diamondicity.

**Question 6.1.** *Is it possible to construct a family of universal partial words over  $A^n$ , where  $n \geq 5$  with diamondicity  $d = 1$ ?*

In addition, we have been unable to find nontrivial examples of universal partial words with diamondicity greater than 1. While such words are shorter, there are many more initial window frames to check.

**Question 6.2.** *Is there a universal partial word over a non-binary alphabet with diamondicity  $d > 1$ ?*

In Section 4, we were able to find an extremal bound on diamondicity for a given  $n$ , but we would like to find a bound which is a constant fraction of  $n$ .

**Question 6.3.** *Is there a  $\varepsilon \in (0, 1)$  such that there does not exist a universal partial word of length  $n$  with diamondicity  $d \geq \varepsilon n$  for all  $n$  sufficiently large?*

Enumerative questions remain largely unstudied.

**Question 6.4.** *For a given  $n$  and  $A$ , how many universal partial words for  $A^n$  exist?*

Chen et al. [5] proved the existence of universal partial words over binary alphabets in several cases via Hamiltonian and Eulerian cycles in de Bruijn graphs. The properties of de Bruijn graphs in higher dimension are less studied, so the proof techniques are not as readily applicable to larger alphabet sizes. Other questions that have been studied in the context of de Bruijn cycles and other universal cycles may also be asked.

**Question 6.5.** *Given a word  $v$  in  $A^n$  and a universal partial word  $w$  for  $A^n$ , how can one efficiently search for  $v$  in  $w$ ?*

## 7. ACKNOWLEDGEMENTS

We thank Jeremy Martin for his role as faculty advisor for the problem and reading over our results. We thank Nathan Graber for his collaboration during the Graduate Research Workshop in Combinatorics. We would also like to thank the organizers of GRWC 2016, as well as the other participants who provided insight on the problem.

## REFERENCES

- [1] Emily Allen, Francine Blanchet-Sadri, Cameron Byrum, Mihai Cucuringu, and Robert Mercas. Counting bordered partial words by critical positions. *the electronic journal of combinatorics*, 18(P138):1, 2011.
- [2] Francine Blanchet-Sadri, Naomi C Brownstein, Andy Kalcic, Justin Palumbo, and Tracy Weyand. Unavoidable sets of partial words. *Theory of Computing Systems*, 45(2):381–406, 2009.
- [3] Francine Blanchet-Sadri, Michelle Cordier, and Rachel Kirsch. Border correlations, lattices, and the subgraph component polynomial. In *Combinatorial Algorithms*, volume 8986 of *Lecture Notes in Comput. Sci.*, pages 62–73. Springer, Cham, 2015.
- [4] Herman ZQ Chen, Sergey Kitaev, Torsten Mütze, and Brian Y Sun. On universal partial words over binary alphabets. *arXiv preprint arXiv:1601.06456v2*, 2016.
- [5] Herman ZQ Chen, Sergey Kitaev, and Brian Y Sun. On universal partial words over binary alphabets. *arXiv preprint arXiv:1601.06456*, 2016.
- [6] Fan Chung, Persi Diaconis, and Ron Graham. Universal cycles for combinatorial structures. *Discrete Mathematics*, 110(1-3):43–59, 1992.